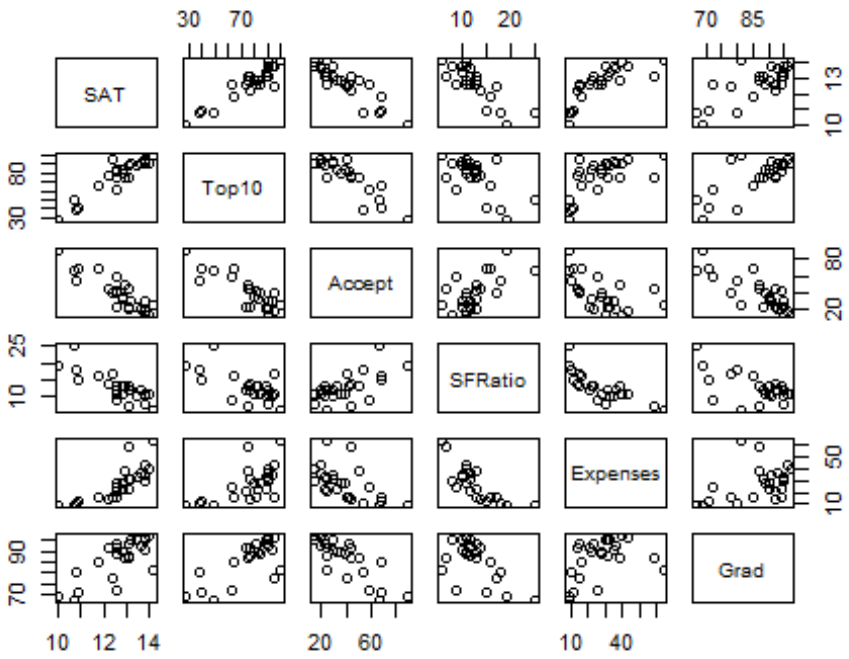
AMS Assignment 5

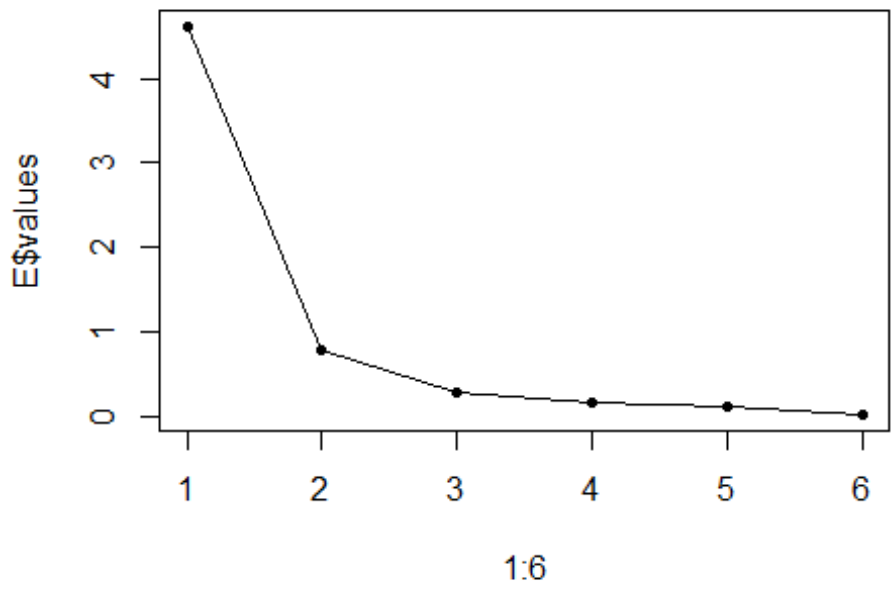
Changrong Xiao

Oct. 15, 2019

```
# Assignment 6
# Exercise 1
Universities = read.table(file = "Universities.txt", header = T, dec = ",")

# a
X = Universities[2:7]
pairs(X)
```




```
Sigma = cor(X)
E = eigen(Sigma)
plot(1:6, E$values, pch = 20)
lines(1:6, E$values)
```



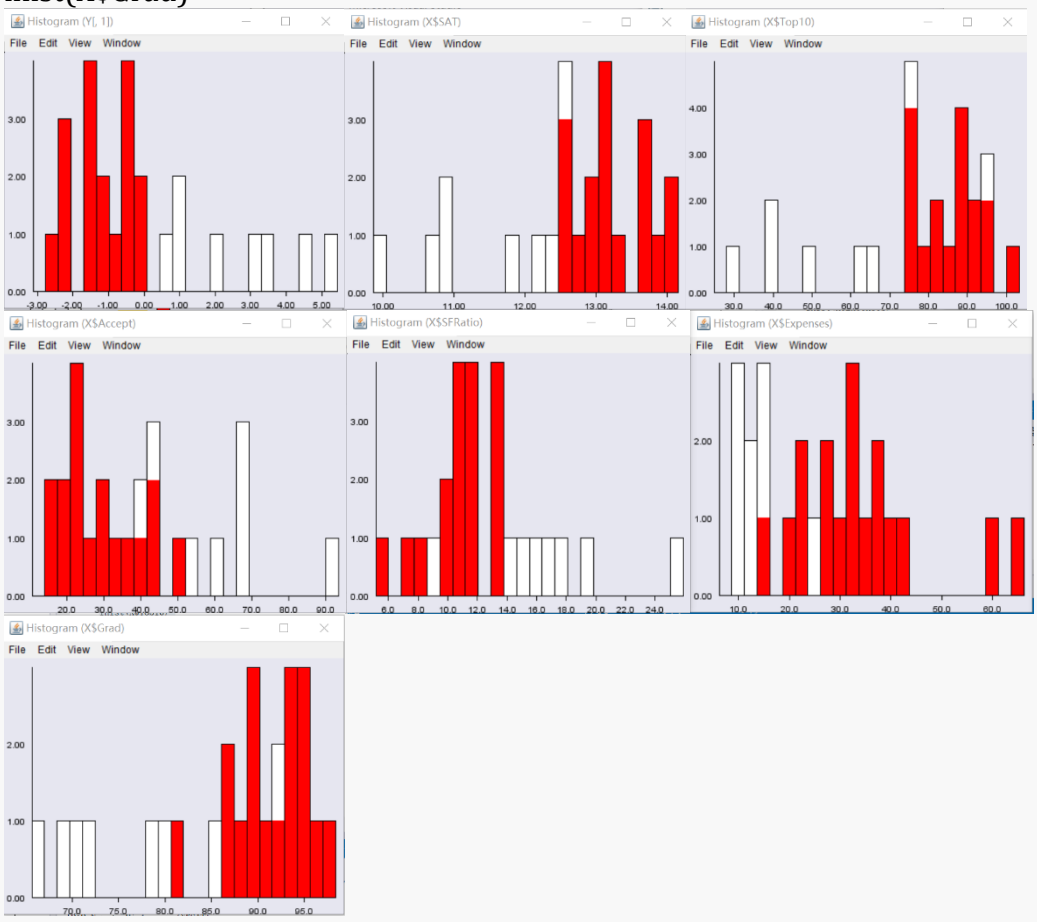

```r
# b
print(round(Sigma, 4))
```

```
##              SAT   Top10  Accept SFRatio Expenses    Grad
## SAT       1.0000  0.9225 -0.8858 -0.8126   0.7790  0.7477
## Top10     0.9225  1.0000 -0.8592 -0.6434   0.6115  0.7459
## Accept   -0.8858 -0.8592  1.0000  0.6317  -0.5584 -0.8195
## SFRatio  -0.8126 -0.6434  0.6317  1.0000  -0.7818 -0.5609
## Expenses  0.7790  0.6115 -0.5584 -0.7818   1.0000  0.3936
## Grad      0.7477  0.7459 -0.8195 -0.5609   0.3936  1.0000
```

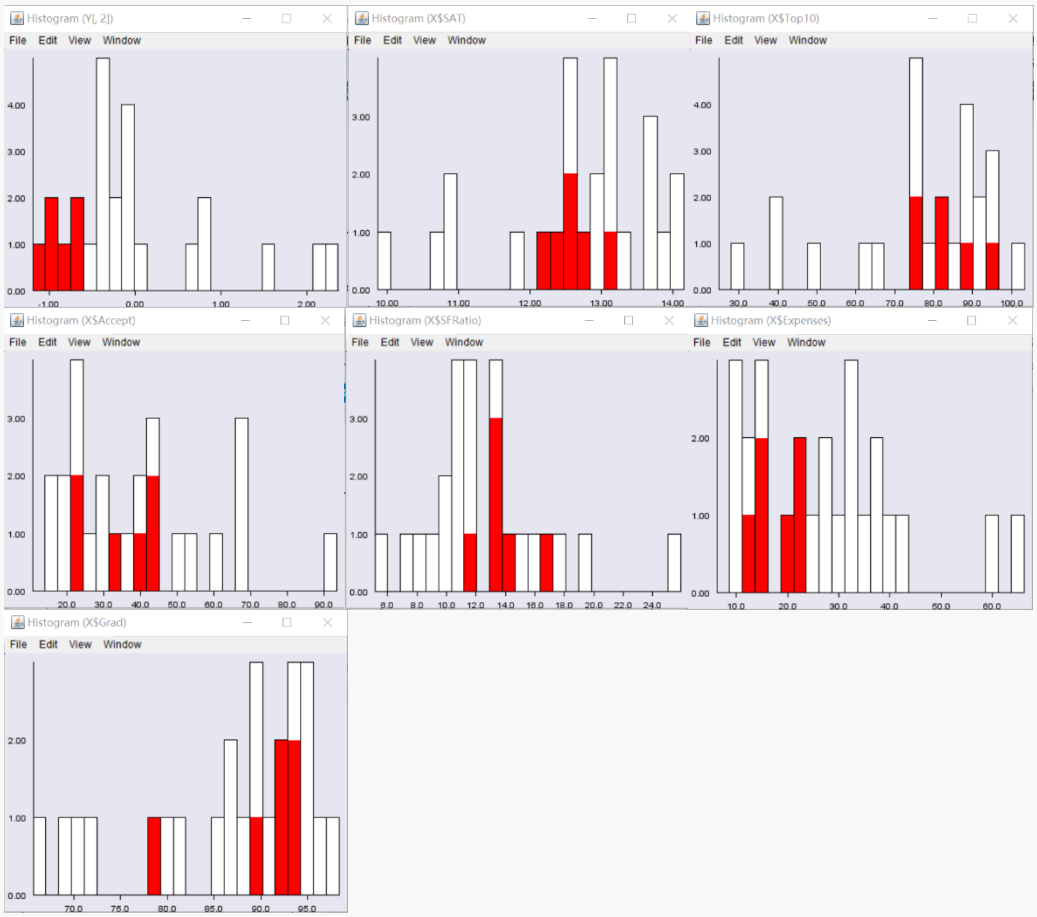
```r
# c
library(iplots)
X_s = scale(X, center = T, scale = T)
Y = X_s %*% E$vectors
```

```
ihist(Y[, 1])

ihist(X$SAT)

ihist(X$Top10)

ihist(X$Accept)

ihist(X$SFRatio)

ihist(X$Expenses)
```

ihist(X$Grad)




```
ihist(Y[, 2])

ihist(X$SAT)

ihist(X$Top10)

ihist(X$Accept)

ihist(X$SFRatio)

ihist(X$Expenses)

ihist(X$Grad)
```

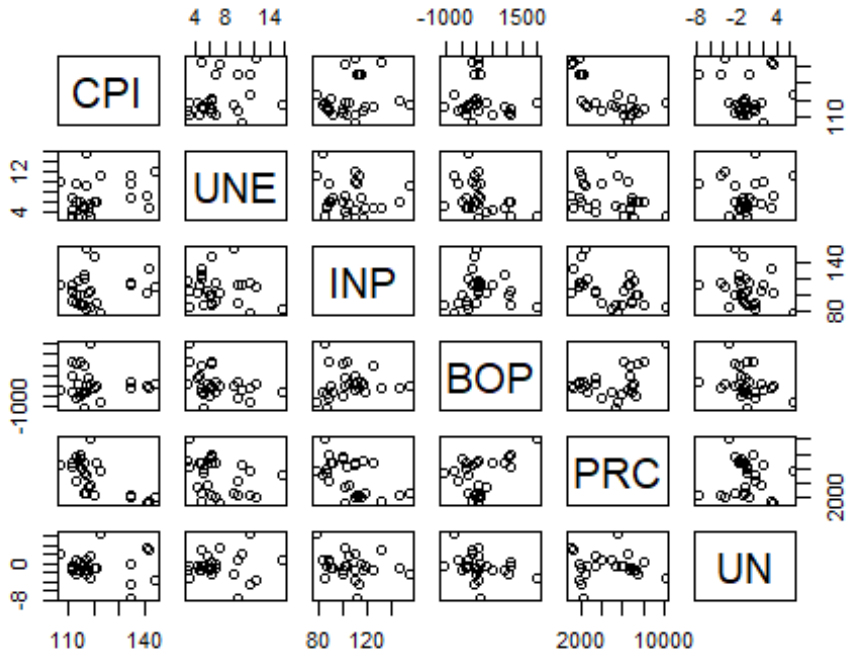



```r
# d
get_pca_table = function(pca, Matrix, size) {
    print(summary(pca, loadings = T))
    E = eigen(Matrix)
    print("The eigenvalues: ")
    print(round(E$values, 4))

    rho = matrix(data = NA, size, size)
    for (i in 1:size) {
        for (j in 1:size) {
            rho[i, j] = E$vectors[i, j] * sqrt(E$values[j]) / sqrt(Matr
ix[i, i])
        }
    }
    print("The correlation coefficients between PCs and variables: ")
    print(round(rho, 4))
}

options(digits = 4)
pca_cor = princomp(covmat = Sigma)
get_pca_table(pca_cor, Sigma, 6)
```

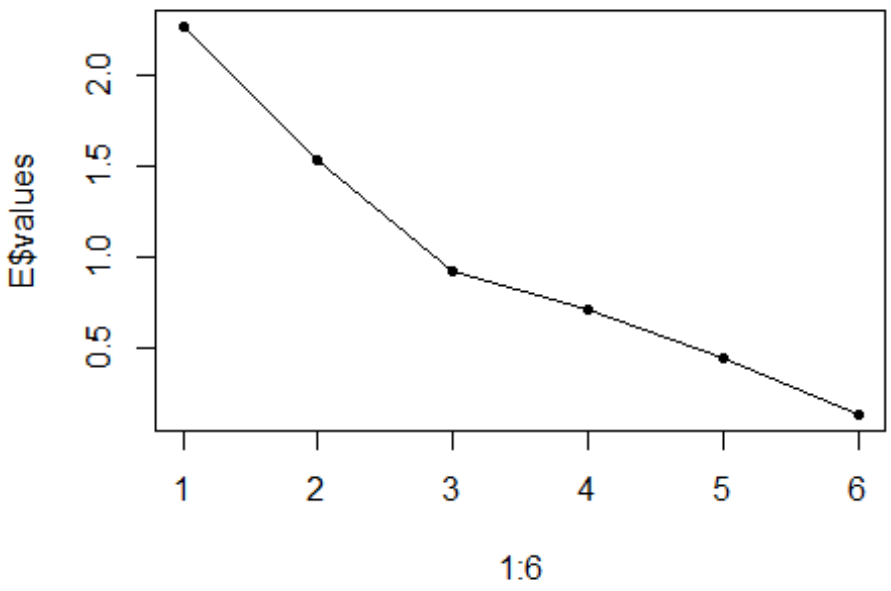
```
## Importance of components:
##                           Comp.1 Comp.2  Comp.3 Comp.4  Comp.5   Comp.6
## Standard deviation     2.1476 0.8870 0.53531 0.4047 0.35257 0.162636
## Proportion of Variance 0.7687 0.1311 0.04776 0.0273 0.02072 0.004408
## Cumulative Proportion  0.7687 0.8998 0.94758 0.9749 0.99559 1.000000
##
## Loadings:
##          Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
## SAT      -0.458        -0.187  0.131         0.858
## Top10    -0.427 -0.200 -0.498  0.375  0.482 -0.396
## Accept    0.424  0.321  0.156         0.801  0.217
## SFRatio   0.391 -0.433 -0.606 -0.507         0.172
## Expenses -0.363  0.634 -0.205 -0.623        -0.174
## Grad     -0.379 -0.516  0.532 -0.439  0.338
## [1] "The eigenvalues: "
## [1] 4.6121 0.7868 0.2866 0.1638 0.1243 0.0265
## [1] "The correlation coefficients between PCs and variables: "
##         [,1]    [,2]    [,3]    [,4]   [,5]    [,6]
## [1,] -0.9831  0.0352 -0.1001  0.0531 0.0073  0.1396
## [2,] -0.9173 -0.1773 -0.2665  0.1517 0.1699 -0.0644
## [3,]  0.9112  0.2846  0.0837  0.0248 0.2824  0.0353
## [4,]  0.8389 -0.3837 -0.3244 -0.2053 0.0271  0.0280
## [5,] -0.7785  0.5628 -0.1096 -0.2523 0.0256 -0.0283
## [6,] -0.8148 -0.4573  0.2850 -0.1775 0.1192 -0.0006

# f
classify = function(Y, Name) {
    category = rep(0, length(Y[, 1]))
    category[which(Y[, 1] > 0 & Y[, 2] > 0)] = 1
    category[which(Y[, 1] < 0 & Y[, 2] > 0)] = 2
    category[which(Y[, 1] < 0 & Y[, 2] < 0)] = 3
    category[which(Y[, 1] > 0 & Y[, 2] < 0)] = 4

    print("PC1 > 0 and PC2 > 0: ")
    print(Name[which(category == 1)])

    print("PC1 < 0 and PC2 > 0: ")
    print(Name[which(category == 2)])

    print("PC1 < 0 and PC2 < 0: ")
    print(Name[which(category == 3)])

    print("PC1 > 0 and PC2 < 0: ")
    print(Name[which(category == 4)])
}

classify(Y, Universities$University)

## [1] "PC1 > 0 and PC2 > 0: "
## [1] CarnegieMellon UWisconsin     Purdue
```

```
## 25 Levels: Brown CalTech CarnegieMellon Columbia Cornell ... Yale
## [1] "PC1 < 0 and PC2 > 0: "
## [1] MIT          CalTech      JohnsHopkins UChicago
## 25 Levels: Brown CalTech CarnegieMellon Columbia Cornell ... Yale
## [1] "PC1 < 0 and PC2 < 0: "
##  [1] Harvard      Princeton    Yale         Stanford     Duke

##  [6] Dartmouth    Brown        UPenn        Cornell      Northwester
n
## [11] Columbia     Georgetown
## 25 Levels: Brown CalTech CarnegieMellon Columbia Cornell ... Yale
## [1] "PC1 > 0 and PC2 < 0: "
## [1] NotreDame  UVirginia  UMichigan  UCBerkeley PennState  TexasA&M

## 25 Levels: Brown CalTech CarnegieMellon Columbia Cornell ... Yale
```

```
# Exercise 2
Europe = read.table(file = "Europe.txt", header = T, dec = ",")

# a
X = Europe[2:7]
pairs(X)
```

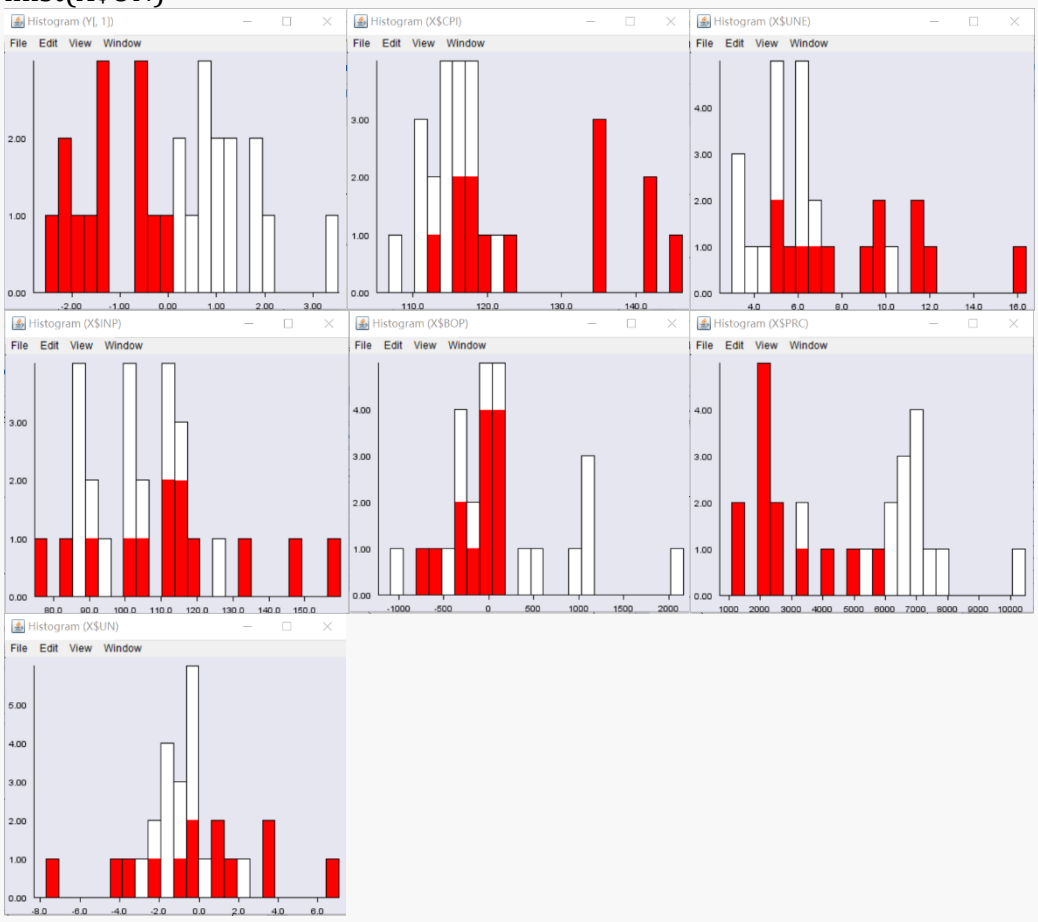



```
Sigma = cor(X)
E = eigen(Sigma)
plot(1:6, E$values, pch = 20)
lines(1:6, E$values)
```




```
# b
print(round(Sigma, 4))

##         CPI     UNE     INP     BOP     PRC      UN
## CPI  1.0000  0.2890  0.2117 -0.1089 -0.7064 -0.0779
## UNE  0.2890  1.0000 -0.1042 -0.4037 -0.3328  0.0255
## INP  0.2117 -0.1042  1.0000  0.0400 -0.5113 -0.2383
## BOP -0.1089 -0.4037  0.0400  1.0000  0.4172 -0.3129
## PRC -0.7064 -0.3328 -0.5113  0.4172  1.0000 -0.0200
## UN  -0.0779  0.0255 -0.2383 -0.3129 -0.0200  1.0000

# c
library(iplots)
X_s = scale(X, center = T, scale = T)
Y = X_s %*% E$vectors

ihist(Y[, 1])

ihist(X$CPI)

ihist(X$UNE)

ihist(X$INP)

ihist(X$BOP)

ihist(X$PRC)
```

ihist(X$UN)



ihist(Y[, 2])

ihist(X$CPI)

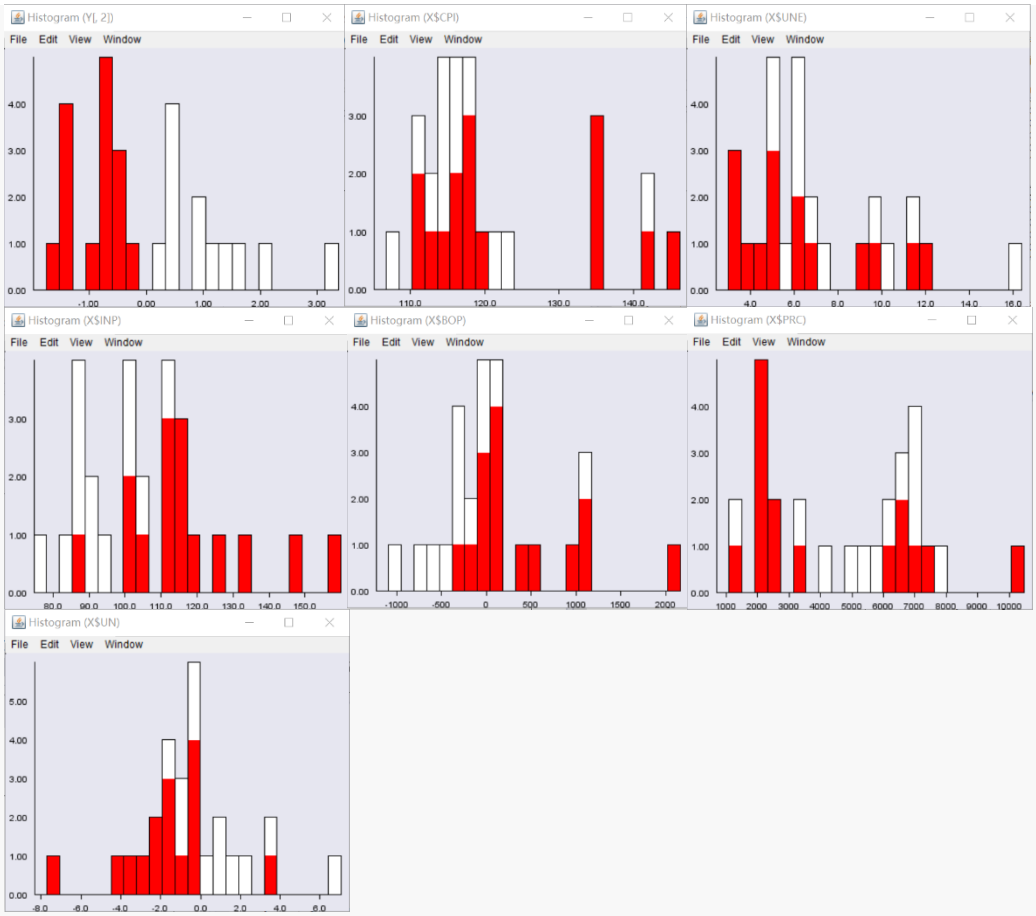
ihist(X$UNE)

ihist(X$INP)

ihist(X$BOP)

ihist(X$PRC)

ihist(X$UN)




```
# d
options(digits = 4)
pca_cor = princomp(covmat = Sigma)
get_pca_table(pca_cor, Sigma, 6)
```

```
## Importance of components:
##                         Comp.1 Comp.2 Comp.3 Comp.4  Comp.5  Comp.6
## Standard deviation      1.5049 1.2392 0.9587 0.8412 0.66466 0.36187
## Proportion of Variance 0.3775 0.2559 0.1532 0.1179 0.07363 0.02183
## Cumulative Proportion  0.3775 0.6334 0.7866 0.9045 0.97817 1.00000
##
## Loadings:
##     Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
## CPI -0.510 -0.170 -0.167  0.644 -0.196 -0.479
## UNE -0.372  0.336 -0.598 -0.213  0.587
## INP -0.290 -0.534  0.432 -0.405  0.390 -0.357
## BOP  0.363 -0.493 -0.168  0.483  0.536  0.277
## PRC  0.620  0.120 -0.175               -0.749
## UN           0.562  0.608  0.373  0.415
## [1] "The eigenvalues: "
## [1] 2.2648 1.5357 0.9192 0.7077 0.4418 0.1310
## [1] "The correlation coefficients between PCs and variables: "
##         [,1]    [,2]    [,3]    [,4]    [,5]    [,6]
```

```
## [1,] -0.7676 -0.2105 -0.1603  0.5420 -0.1303 -0.1733
## [2,] -0.5602  0.4162 -0.5732 -0.1790  0.3900 -0.0146
## [3,] -0.4364 -0.6618  0.4140 -0.3408  0.2594 -0.1293
## [4,]  0.5469 -0.6109 -0.1610  0.4062  0.3561  0.1001
## [5,]  0.9334  0.1490 -0.1682 -0.0483  0.0489 -0.2711
## [6,] -0.0318  0.6962  0.5825  0.3136  0.2761 -0.0208
```

```
# f
classify(Y, Europe$i..Country)
```

```
## [1] "PC1 > 0 and PC2 > 0: "
## [1] Denmark       Ireland       France        Italy         Cyprus

## [6] Finland       UnitedKingdom
## 27 Levels: Austria Belgium Bulgaria Cyprus CzechRep Denmark ... Unit
edKingdom
## [1] "PC1 < 0 and PC2 > 0: "
## [1] Bulgaria Greece   Spain    Portugal Slovenia
## 27 Levels: Austria Belgium Bulgaria Cyprus CzechRep Denmark ... Unit
edKingdom
## [1] "PC1 < 0 and PC2 < 0: "
## [1] CzechRep  Estonia   Latvia    Lithuania Hungary   Poland    Roma
nia
## [8] Slovakia
## 27 Levels: Austria Belgium Bulgaria Cyprus CzechRep Denmark ... Unit
edKingdom
## [1] "PC1 > 0 and PC2 < 0: "
## [1] Belgium     Germany     Luxembourg  Malta       Netherlands Aust
ria
## [7] Sweden
## 27 Levels: Austria Belgium Bulgaria Cyprus CzechRep Denmark ... Unit
edKingdom
```